(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 998 051 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
03.05.2000 Bulletin 2000/18

(51) Int. Cl.⁷: H04B 1/66

(21) Application number: 99121399.2

(22) Date of filing: 27.10.1999

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 29.10.1998 JP 30831298

(71) Applicant:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka (JP)

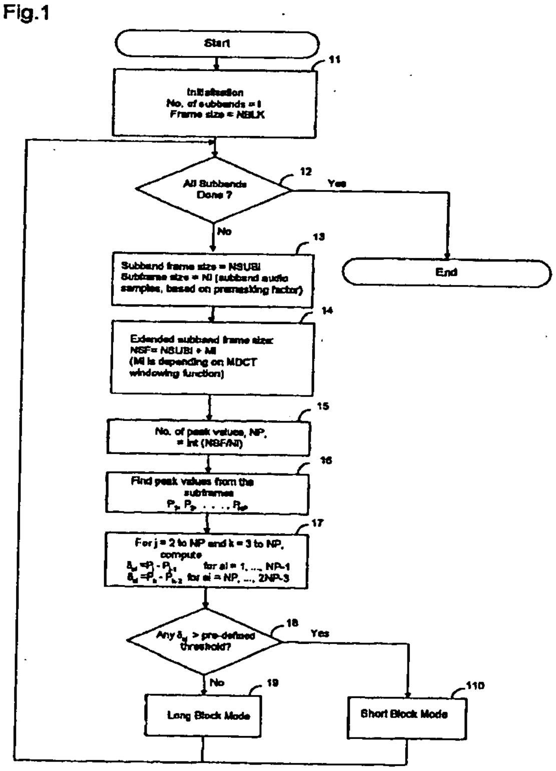
(72) Inventors:
- Shen, Sheng Mei
12-271 Singapore 270007 (SG)
- Neo, Sua Hong
08-239 Singapore 530602 (SG)
- Tan, Ah Peng
10-03 Singapore 550542 (SG)

(74) Representative:
Eisenführ, Speiser & Partner
Martinistrasse 24
28195 Bremen (DE)

# (54) Block size determination and adaptation method for audio transform coding

(57) An effective block size determination methods are disclosed for hybrid coding, especially for ATRAC codec system. They are an improved subframe division method, and peak energy centered method. The accurate detection of an attack signal is very important in a hybrid audio coding system, in order to eliminate or reduce significantly the pre-echo noise. These methods, compared with the prior art, can provide a more accurate block size determination, and have similar level of complexity as the prior art.

Fig.1




EP 0 998 051 A2

## Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to the efficient information coding of digital audio signals for transmission or digital storage media.

2. Description of the Related Art

[0002] Audio compression algorithms using various frequency transforms such as subband coding, adaptive transform coding or their hybrids have been developed and used in a variety of commercial applications. Examples of adaptive transform coders include those reported by K. Brandenburg et al in "Aspec: Adaptive spectral entropy coding of high quality music signals", 90$^{th}$ AES Convention, February 1991 and by M. Iwadare et al in "A 128 kb/s Hi-Fi Audio Codec based on Adaptive Transform Coding with Adaptive Block Size MDCT", IEEE Journal on Selected Areas in Communications, Vol. 10, No. 1, January 1992. Examples of algorithms using hybrid subband and adaptive transform coding include the ISO/IEC 11172-3 Layer 3 algorithm and the ATRAC compression algorithm used in the MiniDisc system. Details of these algorithms can be found in the "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s Part 3: Audio (ISO/IEC 11172-3: 1993)" document and chapter 10 of the MD system description document by Sony in Sep 1992 respectively. The transform filter bank used by these algorithms is typically based on Modified Discrete Cosine Transform, first proposed by Princen and Bradley in "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation", Proceedings of the ICASSP 1987, pp 2161 - 2164.

[0003] In a typical transform encoder, as shown in Fig. 5, the input audio samples are first buffered by buffer 51 in frames, and at the same time passed to a block-size selector 52 to determine the suitable block size or window prior to the windowing and transform by window and transform unit 53, of the audio samples. In a hybrid subband and transform coder such as the ATRAC algorithm, the input audio samples, sampled at 44.1 kHz, i.e., 44100 samples generated per second, are subjected to a hybrid subband and transform coding. The hybrid subband-transform front-end of the encoding process of the ATRAC algorithm is shown in Fig. 6. The input audio samples are first subband filtered into two equal bandwidths using quadrature mirror filter 61 and the resultant lower frequency band is further subdivided into two equal bandwidths by another set of quadrature mirror filter 62. Here, L, M, H means Low band, Middle band, and High band, respectively. A time delay 63 is used to time-align the signal in high-frequency band with those of the lower frequency bands. The subband samples are then separately passed to the block size selector 64 to determine suitable block sizes for the windowing and the modified discrete cosine transform processes in blocks 65, 66 and 67. One of the two block sizes or modes will be selected for each of the frequency bands. The transformed samples are then grouped into units and within each unit, a scale factor equivalent to or just exceeding the maximum amplitude of the unit samples is selected. The transformed samples are then quantized using the determined scale factors and the bit allocation information derived from the dynamic bit allocation unit 68.

[0004] It is known that, in transform coding, a pre-echo or a noise/ringing effect in the silent period before a sudden increase of signal magnitude, or an attack, can occur, particularly if transform coding block size for the audio frame containing the attack is long. Modified discrete cosine transform with adaptive block sizes is typically used to reduce the pre-echo as well as the noise at block boundaries. The block sizes available for the transform coding must in the first place be selected such that if a signal attack were to be detected, a short block transform could be used to process the attack signal and will not give rise to ringing or noise signal to the adjacent blocks. When the size of the short block is made small enough, the pre-echo will not be audible. An important issue is the accurate detection of an attack signal itself.

[0005] The block size decision method outlined in the MD system description of Sep 1992 is shown in Fig. 7. The peak detection step 71 identifies the peak value within each 32 sample block. The adjacent peak values are then compared in step 72. In the decision step 73, where the difference exceed 18 dB, mode 1 or the short block mode step 74 is selected. Otherwise, mode 3 or mode 4 which is the long block mode step 75, for the different frequency bands, will be selected.

[0006] A highly effective audio signal classification and block size determination method is needed for very good reduction of pre-echo during adaptive transform or hybrid subband-transform coding. This is to render the pre-echo to be totally inaudible. While it is recognised that the actual block sizes being used for the transform is in itself an important factor, the accurate detection of signal attack and particularly the critical ones is very significant. Generally, it is desired to use long block for transform coding of the audio signals as the corresponding better frequency resolution obtained will give rise to more accurate redundancy and irrelevancy removal of the audio signal components. This is especially true for segments of the audio signals where the characteristics of the audio signal varies slowly. Short blocks are to be used only when identified to be absolutely necessary and for the critical attack signals. The block size decision method provided in the prior art does not give good result in transient or attack signal detection accuracy. It can fail to detect an attack signal

which occur within a time interval of premasking duration. Premasking is the condition whereby presence of a fast buildup of loud sounds or attack occurring in time has a masking effect on the sound preceding the attack. Failure of such detection can sometimes give rise to undesirable audible effects. While single-tone masker experiments have demonstrated premasking duration lasting between 5 ms and 20 ms, empirically, pre-echo of shorter duration has been audible. The effective premasking duration is expected to be in the region of less than 5 ms. Post masking effect, the lingering masking effect after occurrence of a masker, typically spans 20 ms or more. Where long block frame size is typically less than 20 ms, the release of a peak signal is normally regarded as having insignificant effect. For very high accuracy block size determination, post masking effect could be taken into account.

## SUMMARY OF THE INVENTION

**[0007]** This invention is based on the need for a high accuracy block size decision scheme and has taken into account temporal masking considerations, both the premasking and postmasking effects. In this invention, means of operating on full bandwidth audio signals or on limited bandwidth signals, for example, after subband filtering into frequency bands are possible. This invention has the means of grouping audio samples in a current considered frame into subframes of equal time interval of approximately 3 ms, in consideration of empirical premasking duration, excepting the final subframe which is of half the time interval; this said current considered frame together with the whole or half of the final subframe of the previous considered frame, and optionally, half subframe from the future frame constituting the extended frame, will be used for peak value estimate; means to identify the said peak values within the said subframes; means to compute the differences between said peak values of adjacent time intervals; means to, optionally, compute the differences between said peak values separated by a subframe time interval; means to decide whether long block size or short block size should be used after comparing the said differences with predetermined threshold. An alternative method comprises the means of grouping samples in current frame together with the whole or half of the final subframe of the previous considered frame, the said subframe interval being determined by the temporal hearing characteristics of the human ear; means of identification of a selected number of peak values within the resultant grouping; means of the designating a peak value, selected in order of magnitude, as reference peak value; means of identification of the peak value from within a subframe interval preceding the reference peak value; means of computation of the difference between the reference peak value and the peak value within the said subframe interval preceding it; means of comparison of the said difference with predetermined threshold values, wherein a smaller block size is invoked when the difference exceed the predetermined threshold value; otherwise a new reference peak is used and the process repeated until a difference exceeding the predetermined threshold is found or when all the available peak values have been considered.

**[0008]** The means of grouping audio samples in a current considered frame into subframes, first involve taking a designated number of audio samples from the previous frame and optionally the future frame, together with all samples in the current frame. The time interval for each subframe should span approximately 3 ms, based on an empirically determined premasking duration. The designated number of audio samples should be approximately of half subframe duration. The grouping into subframes can proceed as illustrated in Fig. 3. In allowing for computing the difference between the peak of a current subframe and up to two previous subframes will allow for a wider scope of signals to be classified as attack signals. The obtained difference in peak values is then compared against a positively set threshold value. This means that the post masking effects of release of signal will be ignored. Should it be desired to consider the effects of the less significant postmasking, comparison against a negative threshold will be necessary. The first set of means whereby the audio samples are first grouped into subframes provide for a convenient and computationally less intensive method of obtaining peak values and computation of differences for the purpose of block size determination. However, the set of means does not thoroughly search for all possible signal attacks or transients. The alternative set of means whereby a selected number of peak values are first identified within the said extended frame will allow a more thorough search. Subject to computational load permitted, a maximum number of peaks is first identified. The highest peak value is first taken as the reference peak. From a time window of a subframe from this reference peak, a peak value is established and the difference with the reference peak is computed. If the difference is not larger than predetermined threshold; the procedure is repeated using the second highest peak value as reference peak value and so forth. The process is repeated until a difference exceeding the predetermined threshold is found or when all the available peak values have been considered.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0009]**

Fig. 1 is the flow chart of an embodiment of the invention of an improved subframe division block size determination method.

Fig. 2 is the flow chart of a second embodiment of the invention of peak energy centered block size determination method.

Fig. 3 is an illustration of a subframe method and the difference computation.

Fig. 4 is an illustration of a peak energy centered block size determination method.

Fig. 5 is a block diagram of the front-end of an adaptive transform encoder.

Fig. 6 is a block diagram of the front-end of the ATRAC encoder.

Fig. 7 is a flow chart of the prior art of the block size determination method.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

**[0010]** A flow chart of an embodiment, termed as Improved Subframe Division Block Size Determination Method, is shown in Fig. 1. In the general context, a hybrid subband transform coder is inherently assumed. In the case where only transform coding is purely used, the number of subband may be treated as 1. Each subband frame as defined in step 14 of Fig. 1 is partitioned into subframes. For the purpose of illustration, an example using a subband frame size of 128 samples is shown in Fig. 3. A subband subframe size of 32 samples which translates to a time interval of approximately 3.0 ms is adequate based on premasking considerations.

**[0011]** In this embodiment, there are two major differences from the prior art. One is the 16-sample extension of the current subband frame size of 128 for detecting an attack signal, instead of only using 128 samples. This extension comes from the windowing function of MDCT. The other is to check the difference between peak values separated by a subframe segment, besides to check the difference between the adjacent peak values, if the difference between adjacent peak values is less than the predetermined threshold. Both are required to reduce probability of miss in the detection of attack signal.

**[0012]** After the initialization of the number of subbands and frame size in step 11, block size determination is performed for each and every subband. A decision step 12 ascertains whether all the subbands have been analyzed. Depending on the types of subband filtering performed, whether equal or unequal subband bandwidths are used for all subbands, the value assignment of subband frame size and the appropriate subframe size in step 13 will vary accordingly. In step 14, each subband frame is extended to NSF (=NSUBi + Mi) by taking into account for all covered samples by window function of MDCT. Here, Mi is the number of the extended samples. For example, for MDCT of 32-sample, the number of extended samples is 16.

**[0013]** The number of segments for the purpose of peak identification is computed in step 15. The peak values within each segment are identified in step 16. The differences between adjacent peak values and between peak values separated by a subframe segment are computed in step 17. As long as a single difference exceed thepredetermined threshold, as determined in the decision step 18, a short block assignment step 110 will be performed. Otherwise, a long block assignment step 19 will be provided.

**[0014]** The extended subband frame, as illustrated in Fig. 3, is formed for the purpose of peak value identifications. Based on the example of Fig. 3, where 32 is the subframe size, the number of the extended samples will be 16 based on the window function used for MDCT of 32-sample. The subband frame size of 128 samples together with the 16 samples from the previous frame will be considered for attack signal detection. Therefore, four 32-samples sub-frames and one 16-sample subframe will be used in each determination iteration. Here the 16 samples which come from the future frame, as shown in Fig. 3, can be neglected since the windowing values drop sharply in this period and also this part is the final part of the extended subband frame. So, the number of peak values to be computed is 5. Altogether, a maximum of 7 difference computations among the peak values will be performed. For implementation efficiency, as soon as one computed difference exceed the predefined threshold, the short block mode will be activated. Typically, the additional comparisons between P3 and P1, P4 and P2, P5 and P3 are needed when all of $\delta_i$ (i=1,2,3,or 4) are less than the predetermined threshold. As long as one of $\delta_i$ (i=1 ,2,3,4,5,6,or7) is larger than the predetermined threshold, then the comparison may be stopped to save computation time.

**[0015]** An alternative embodiment, termed as Peak Energy Centered Block Size Determination Method, is shown in Fig. 2. An attack signal may be considered as the energy of the signal rising sharply over certain duration of a signal. Approximately, the instance of the peak value in a signal may be regarded as the center of the sharply rising energy if there is an attack signal in the same duration, as shown in Fig. 4. This is true for many instances, by empirical observation.

**[0016]** As shown in Fig. 4, P is the peak value of the signal in the period of SD. C is the position of the peak value of P, and it is the focal point of the energy of the signal in the period of SD. The point B is just 32 samples from point C. Another peak value is searched starting from point B to point A, treating it as a 32-sample subblock. If the peak value P is larger than the second peak value, Ps, compared to the predetermined threshold, then it is determined that there is an attack signal appearing in this current block. Short block MDCT will then be applied to this current block. Otherwise, the second peak value Ps is taken as the new P, and the above steps are then iteratively applied until the point S is reached. If there is no other peak value, larger than its second peak value Ps then long block MDCT is applied.

**[0017]** Fig. 2 is the flow chart for peak energy centered block size determination method. The meaning of P, Ps, C, B, A, and S is as shown in Fig. 4. In Fig. 2, step 21 is the initialization of the block size determination for

an audio signal. Step 22, is to check whether all the subbands have been examined for block size determination. If "Yes", the process will then be terminated. Otherwise block size determination will be performed for the following subbands.

[0018] In step 23, the peak value P is found for the current subband frame, and the Peak Energy Centred point C is correspondingly located. In step 24, the rising envelope period of peak energy value P is assumed to be the BC segment which is 32-sample subblock, started from the center point of C. In step 25, the second peak value Ps is searched in a 32-sample of subblock, bounded by points B and A. If the second peak value Ps is less than P by a predetermined threshold in step 26, then a short block mode will be assigned in step 27. Otherwise, in step 28, a check of whether point A coincides with the beginning point of the subband S is done. If not, then Ps is taken as the new P, in step 210 and the above steps 24, 25, 26, 27, and 28 are repeated. If "YES", then a long block mode will be assigned for the current subband frame.

[0019] Fig. 4 is valid when the length of SC is longer than two times of 32-sample. For cases when the length of SC is shorter than two times of 32-sample, the length of BC is not fixed at 32, but will be SC/2, which is less than 32-sample. For this case, the length of SB is also less than 32, and will be SC/2. That is to say, B will be the middle point between S and C.

[0020] In the case when the length of SC is shorter than 16-sample, half of the 32-sample, which is corresponding to 1.45ms, a long block MDCT will be used to the current block. Even in this case when an attack signal in the beginning part of the current block exists, premasking can mask out the short period of pre-echo of less than 29 ms, which is caused by this attack signal.

[0021] The present invention is highly effective in the detection of audio signal attacks, and optionally, the release of the signal. The use of any of the described block size decision techniques will result in highly accurate detection of the critical transient signal attacks, consequentially leading to reduction or elimination of audible pre-echo. This is possible provided appropriate block sizes for the transform coding are used. Different technique options are incorporated, depending on the amount of computational load, RAM and ROM supportable.

## Claims

1. A method of identification and categorisation of an audio signal into subclasses to determine the subframe block size of a transform coder, said method comprising:

   a) detecting the number of block sizes available for the transform coder;
   b) sampling an input audio signal at time intervals into samples and grouping said samples into frames each having an equal number of samples;
   c) analyzing said frames in time domain to produce at least one comparison index;
   d) selecting an appropriate block size for the transform coder.

2. A method according to claim 1, wherein said audio signal is a full bandwidth audio signal.

3. A method according to claim 1, wherein said audio signal is a limited bandwidth audio signal.

4. A method according to claim 1, wherein said analyzing steps comprises:

   a) extending each said frame in accordance with the window function used in said transform coder;
   b) subdividing said extended frame containing the audio samples into smaller subframes, the number of smaller subframes being determined by time interval determined by temporal hearing characteristics of the human ear;
   c) identifying a peak value within each said subframe based on the amplitude of the samples within said subframe;
   d) computing the difference between the peak value of adjacent subframes and the peak value of two subframes which are separated by a subframe time interval, said difference being used as said comparison index;
   e) comparing said index with a predetermined threshold value, such that a smaller block size is invoked when the index is greater than the predetermined threshold value, and a larger block size is invoked when the index is not greater than the predetermined threshold value.

5. A method according to claim 1, wherein said analyzing step comprises:

   a) extending each said frame/subband frame by considering the window function used in the said transform coder;
   b) identifying a designated number of peak values within each said extended frame /extended subband frame based on the amplitude of the samples within the extended frame, each peak value being a local maximum amplitude value;
   c) identifying the subframe interval as determined by the temporal hearing characteristics of the human ear, taking the highest of the said peak values as reference peak value, identifying the peak value from within a subframe interval preceding the reference peak value;
   d) computing the difference between the refer-

ence peak value and the peak value within said subframe interval preceding it; and

e) comparing said difference with predetermined threshold values, wherein a smaller block size (or subframe size) is invoked when the difference exceed the predetermined threshold value.

6. A method according to claim 5, wherein said comparing step comprises:

a) repeating said steps c) to e) of claim 5 when the difference in step e) of claim 5 does not exceed the predetermined threshold value, by taking the peak value found in step c) of claim 5 as the new reference peak value; and

b) determining a large block size (or subframe size) when no difference exceeding the predetermined threshold value can be found after all the local maximum values have been exhausted in the last subframe interval.

7. A method according to claim 6, wherein said last subframe interval may be equal to or less than its previous subframe interval, which depends on the real situation of each extended frame/subband frame;

8. A method according to claim 5, wherein identification of the peak value from between two subframe interval and one subframe interval preceding the reference peak value take place when step e) of claim 5 does not yield a difference exceeding predetermined thresholds.

9. A method according to claim 4, wherein said audio samples within the final said subframe interval of the preceding audio frame are taken into account for computation of said difference between peak values.

10. A method according to claim 5, wherein said audio samples within the final said subframe interval of the preceding audio frame are taken into account for computation of said difference between peak values.

11. A method of identification and categorisation of audio signal into subclasses to determine the block size (or subframe size) of a transform coder, said method comprising:

a) partitioning the audio signal into different frequency bands;

b) grouping the audio samples in each and every frequency band into frames of equal time interval, the number of said audio samples in frames belonging to different frequency bands may not necessarily be equal;

c) subjecting each said frame of equal interval to an analyzing method giving rise to different block size (or subframe size) decisions for different frequency bands.

12. A method of claim 11, wherein said analyzing method comprises

a) extending each said frame in accordance with the window function used in said transform coder;

b) subdividing said extended frame containing the audio samples into smaller subframes, the number of smaller subframes being determined by time interval determined by temporal hearing characteristics of the human ear;

c) identifying a peak value within each said subframe based on the amplitude of the samples within said subframe;

d) computing the difference between the peak value of adjacent subframes and the peak value of two subframes which are separated by a subframe time interval, said difference being used as said comparison index;

e) comparing said index with a predetermined threshold value, such that a smaller block size is invoked when the index is greater than the predetermined threshold value, and a larger block size is invoked when the index is not greater than the predetermined threshold value.

13. A method according to claim 11, wherein said analyzing method comprises:

a) extending each said frame/subband frame by considering the window function used in the said transform coder;

b) identifying a designated number of peak values within each said extended frame /extended subband frame based on the amplitude of the samples within the extended frame, each peak value being a local maximum amplitude value;

c) identifying the subframe interval as determined by the temporal hearing characteristics of the human ear, taking the highest of the said peak values as reference peak value, identifying the peak value from within a subframe interval preceding the reference peak value;

d) computing the difference between the reference peak value and the peak value within said subframe interval preceding it; and

e) comparing said difference with predetermined threshold values, wherein a smaller block size (or subframe size ) is invoked when the difference exceed the predetermined threshold value.

Fig.1

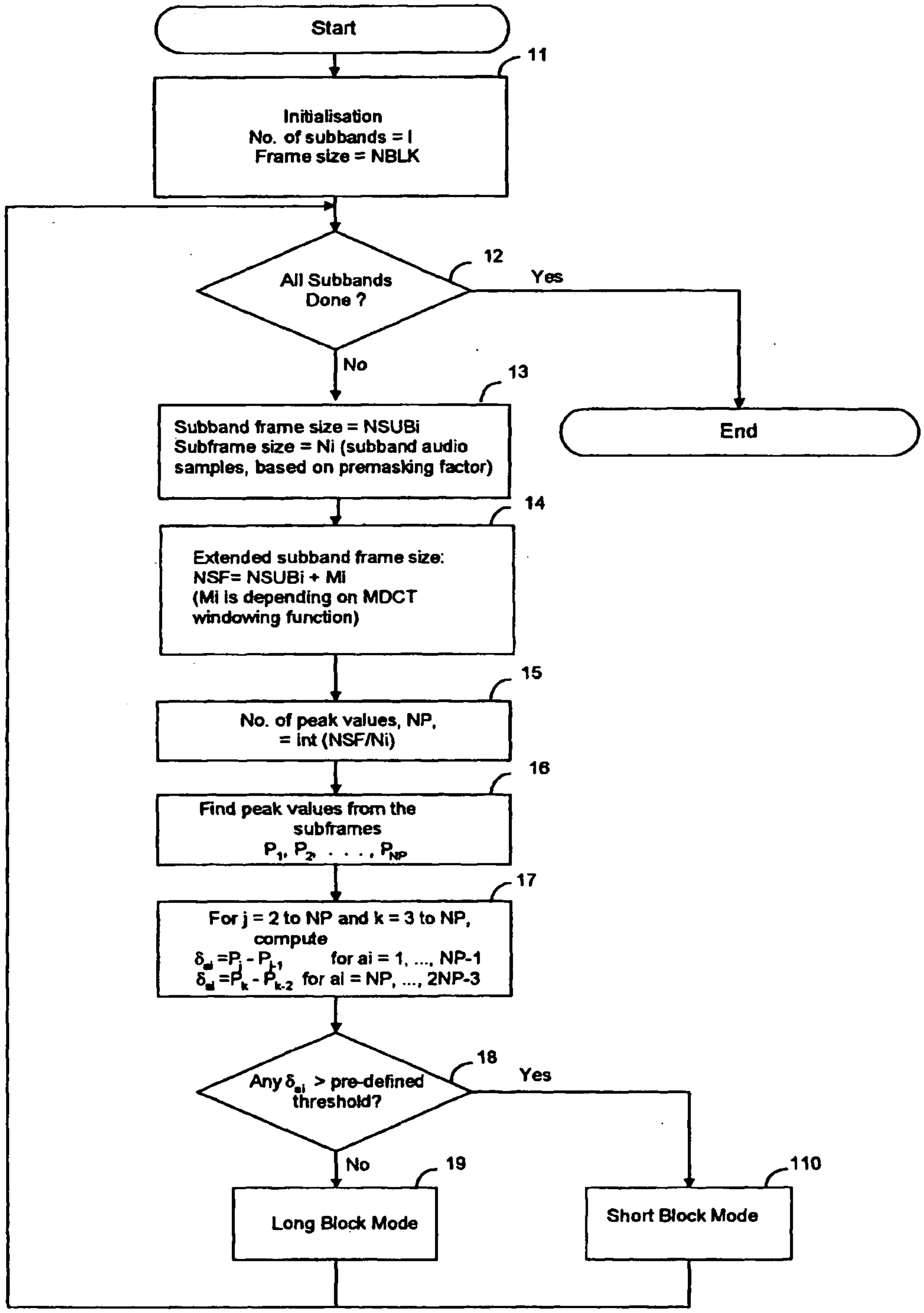

Start
11
Initialisation
No. of subbands = I
Frame size = NBLK
12
All Subbands
Done ?
Yes
No
End
13
Subband frame size = NSUBi
Subframe size = Ni (subband audio samples, based on premasking factor)
14
Extended subband frame size:
NSF= NSUBi + MI
(Mi is depending on MDCT windowing function)
15
No. of peak values, NP, = int (NSF/Ni)
16
Find peak values from the subframes
$P_1, P_2, \ldots, P_{NP}$
17
For j = 2 to NP and k = 3 to NP, compute
$\delta_{ai} = P_j - P_{j-1}$ for ai = 1, ..., NP-1
$\delta_{ai} = P_k - P_{k-2}$ for ai = NP, ..., 2NP-3
18
Any $\delta_{ai}$ > pre-defined threshold?
Yes
No
19
Long Block Mode
110
Short Block Mode

Fig. 2

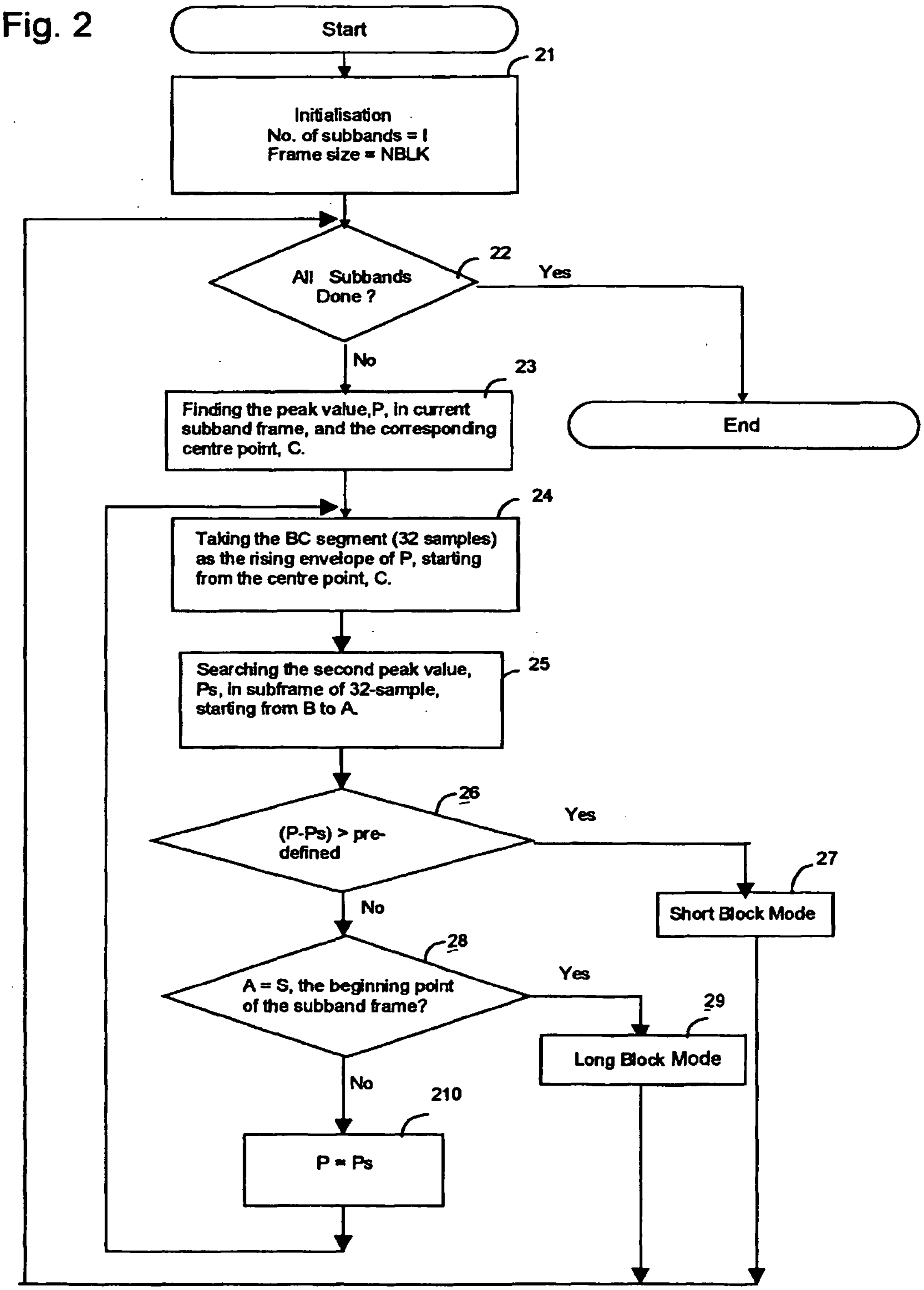

Start
21
Initialisation
No. of subbands = I
Frame size = NBLK
22
All Subbands Done ?
Yes
No
23
Finding the peak value,P, in current subband frame, and the corresponding centre point, C.
End
24
Taking the BC segment (32 samples) as the rising envelope of P, starting from the centre point, C.
25
Searching the second peak value, Ps, in subframe of 32-sample, starting from B to A.
26
(P-Ps) > pre-defined
Yes
27
Short Block Mode
No
28
A = S, the beginning point of the subband frame?
Yes
29
Long Block Mode
No
210
P = Ps

Fig. 3

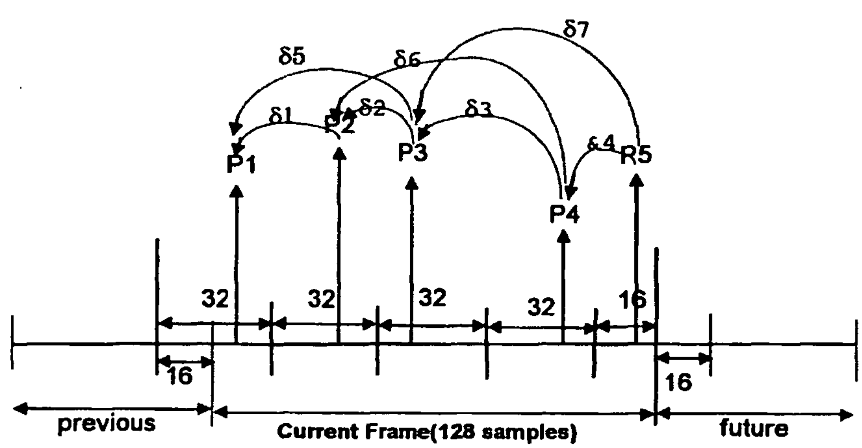

δ7
δ5
δ6
δ1
P2
δ2
δ3
P1
P3
δ4
P5
P4
32
32
32
32
16
16
16
previous
Current Frame(128 samples)
future

Fig. 4

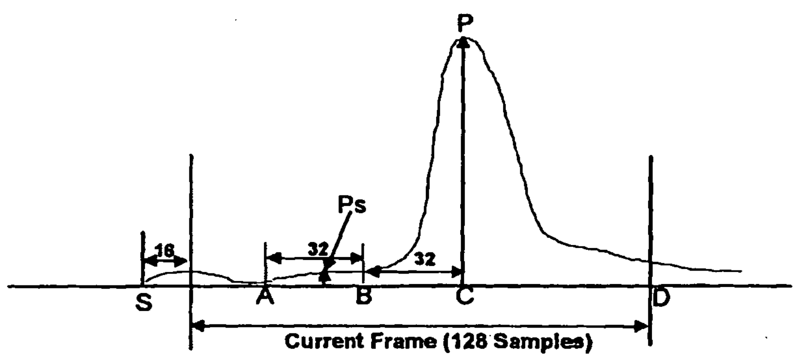


Fig. 5

PRIOR ART

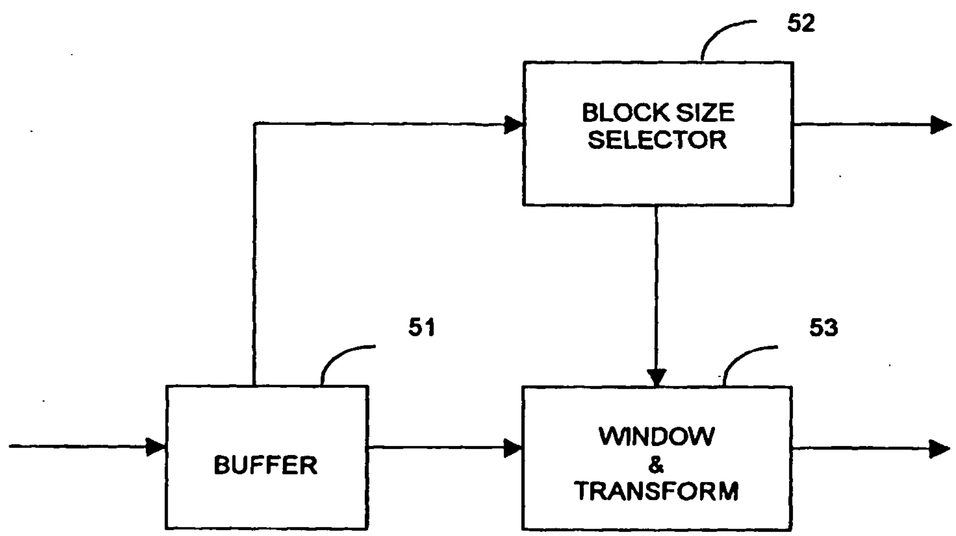


Fig. 6

PRIOR ART

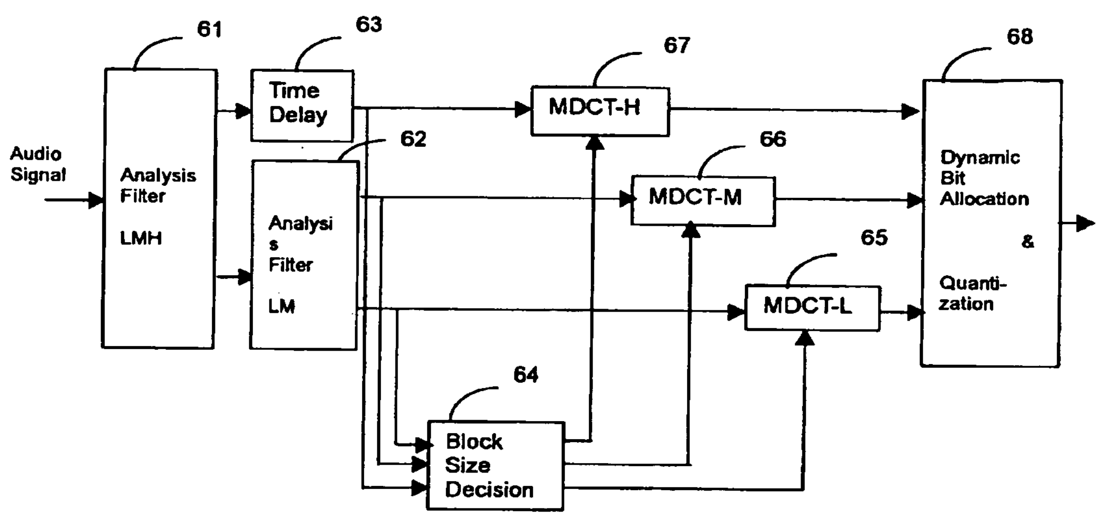


Fig. 7

PRIOR ART